

网络 与 光盘

(化学化工、医药卫生)

科技文献检索

胡先明 周彩霞 罗嘉惠 编著

胡先明 主 编



武汉大学出版社

网络与光盘科技文献检索

(化学化工、医药卫生)

主编：胡先明

编著：胡先明 周彩霞 罗嘉惠

武汉大学出版社

图书在版编目(CIP)数据

网络与光盘科技文献检索:化学化工、医药卫生/胡先明主编;胡先明,周彩霞,罗嘉惠编著. —武汉:武汉大学出版社,2000.11

ISBN 7-307-03031-4

I. 网… II. ①胡… ②胡… ③周… ④罗… III. 计算机网络—计算机应用—科技情报—情报检索 IV. G354

中国版本图书馆 CIP 数据核字(2000)第 33038 号

责任编辑:陈 刚 责任校对:黄添生 版式设计:支 笛

出版:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.whu.edu.cn)

发行:新华书店湖北发行所

印刷:湖北省孝感日报社印刷厂

开本:787×1092 1/16 印张:17.375 字数:427千字

版次:2000年11月第1版 2000年11月第1次印刷

ISBN 7-307-03031-4/G·513 定价:22.00元

版权所有,不得翻印;凡购买我社的图书,如有缺页、倒页、脱页等质量问题者,请与当地图书销售部门联系调换。

内 容 简 介

本书是化学工作者根据自己教学和科研实践编写而成的计算机应用书籍，旨在介绍在计算机网络与光盘上的化学化工、医药卫生文献资料的实用查阅方法。全书共 15 章，分为计算机情报检索概述、国际联机检索系统、各类有关化学化工及医药卫生光盘系统、专利和标准、Internet 国内外网上资源等部分。还介绍了利用 E-mail 检索有关资料文献的方法。在附录中列入了数百个免费发布信息的网址和中英文搜索引擎等内容。

本书可作为高校相关专业的学习教材，是从事化学、医药卫生、生物、环保及材料研究工作人员进行计算机信息查询的必备使用手册，也是广大计算机爱好者上网检索的工具书。

前 言

20 世纪 90 年代以来, Internet 网上资源和光盘数据库得到空前的发展, 广大情报人员、科技工作者急需尽快熟悉和掌握计算机信息检索技术, 改变单一手检方式为与计算机检索相结合的方式。面对网上繁多的信息资源, 需要有一本书告诉我们进入信息高速公路的捷径。本书是作者根据近年来教学与科研实践, 结合有关参考书目, 自己动手编写的一本旨在介绍计算机网络与光盘上的科技文献资料查阅方法的计算机应用书籍。

本书第一章主要介绍计算机情报检索的基本原理和方法; 第二至四章主要介绍当今世界上几个著名的联机检索系统, 是按照从原理到方法、再到实例逐层展开的; 第五至七章选择 CA (Chemical Abstracts) 和 CCWIN (Current Contents) 等为代表, 向读者具体介绍光盘检索系统的有关知识和国内外大型光盘数据库实例; 第八至十章针对专利和标准, 从光盘到网络全面予以介绍, 重点介绍了国内外网络上免费检索专利的网址和相关的检索方法; 第十一至十五章收集了国内外 Internet 上大量可用资源, 包括化学化工、医药卫生及相关领域的多学科网站, 并介绍了使用 E-mail 上网查阅化学化工文献资料的方法; 附录部分列出了数百个免费发布信息网址和中英文搜索引擎。本书不仅对化学化工、医药卫生工作者有帮助, 还可以满足社会各界在这些方面的需求。

本书第四章《OCLC 的网上信息查询系统》、第六章《美国化学文摘光盘数据库》和第八章《网络系统专利文献检索》由罗嘉惠完成; 其他部分由胡先明、周彩霞完成; 最后由胡先明进行全面审核。参加部分编写工作的还有武汉大学测试中心的文健。

在本书编写过程中, 参考了情报检索专家们编写的各种著作和论文, 并得到武汉大学图书馆王梅、湖北省科技情报所吴连荣等同志的帮助和支持, 在此向他们一并表示感谢。

由于各种条件限制和计算机科学技术的迅速发展, 本书的缺憾和错误在所难免, 敬请读者批评指正。

编著者

1999 年 6 月

目 录

第一章 概 述	1
第一节 计算机情报检索的定义	1
第二节 机检系统的主要检索功能	2
第三节 检索策略	5
第二章 DIALOG 查询系统	9
第一节 DIALOG 系统简介	9
第二节 检索方法	10
第三节 检索项	14
第四节 检索逻辑	18
第五节 DIALOG 系统指令	22
第六节 检索输出	43
第七节 DIALOG 系统功能范围	45
第八节 DIALOG 系统标点、符号和检索词	46
第九节 DIALOG 系统的检索举例	48
第三章 STN 系 统	54
第一节 基本检索指令	55
第二节 检索策略的编制	56
第三节 STN 系统数据库的文献记录结构	56
第四节 STN 系统的布尔逻辑算符	58
第五节 位置逻辑符	59
第六节 截词符和屏蔽符	60
第七节 STN 系统的检索指令	61
第八节 STN 联机检索实例	71
第四章 OCLC 的网上信息查询系统	73
第一节 OCLC 信息检索系统	73
第二节 OCLC-FirstSearch 的数据库	74
第三节 OCLC-FirstSearch 的联入方法和检索步骤	78
第四节 FirstSearch Web 的检索方法和检索式构造	82

第五节 检索实例	85
第六节 其他系统	86
第五章 Current Content 检索系统	88
第一节 覆盖范围	89
第二节 检索方法初步	89
第三节 浏览 (Browse)	91
第四节 查找 (Search)	93
第五节 简介编制	97
第六节 查阅 (View)	97
第七节 打印存储	99
第六章 美国化学文摘光盘数据库	102
第一节 DOS 环境下 12 卷累积版	102
第二节 Windows 环境下的 CA 光盘检索系统	113
第七章 常用光盘	119
第一节 美国《博士论文库》检索指南 (DAO Dissertation Abstracts Online)	119
第二节 《研究生指南库》检索指南 (Peterson's Gradline)	122
第三节 《生命科学库》检索指南 (Life Science Collection)	125
第四节 WILSON 光盘检索系统——《生物农业》光盘	128
第五节 《中国学术期刊》全文数据库	135
第六节 中文科技期刊编名库	138
第八章 网络系统专利文献检索	141
第九章 专利文献的光盘检索	153
第一节 中国专利文献的光盘检索	153
第二节 国外专利文献光盘检索	183
第十章 标准信息检索	188
第十一章 中国教育与科研网 (CERNET)	190
第一节 中国教育与科研网简介	190
第二节 CERNET (中国教育与科研网) 资源分布	191
第十二章 中国科学院网络导航	198

第十三章 国内外化学化工网络资源	202
第一节 中国国内化学化工网络资源	202
第二节 国外化学化工网络资源	212
第十四章 医药、保健、卫生国内外网址	238
第十五章 E-mail 的活用	249
第一节 用 E-mail 实现 FTP	249
第二节 用 E-mail 实现 ARCHIE 功能	251
第三节 用 E-mail 实现 GOPHER 功能	251
第四节 用 E-mail 实现 Veronica 功能	252
第五节 用 E-mail 实现 USENET	253
第六节 用 E-mail 实现 WAIS 检索	254
第七节 用 E-mail 实现 WWW	255
第八节 利用 E-mail 查找 WEB 网点	255
第九节 如何用 E-mail 查找医学信息资料	256
附 录	257
一、免费发布信息网址	257
二、中、英文搜索引擎	263
参考书目	269

第一章 概 述

第一节 计算机情报检索的定义

计算机情报检索是利用电子计算机存储和检索情报文献。它包括情报存储和情报检索两大部分。情报存储过程是将原始文献进行主题概念分析、加工处理并以情报检索语言进行文献特征标识后,以数据的形式存入计算机的外部存储介质(如磁带、磁盘、光盘)中,制成文献数据库(database),又称文献资料档,简称文档(file)。情报检索过程是根据用户的情报需求进行主题概念分析,按一定的情报检索语言进行检索提问标识,编成计算机可识别的检索提问式,输入计算机,由计算机将检索提问式与数据库中的文献特征标识进行类比匹配,将符合提问要求的命中文献按用户要求的格式显示、并打印其检索结果。

计算机情报检索的发展过程大致分为如下六个阶段:

(1) 脱机检索过程 计算机用于文献检索开始于 1954 年,美国海军兵器中心首先采用 IBM-701 型计算机建立了世界上第一个科技情报检索系统,但检索结果只是文献号码。1958 年美国通用电器公司将该系统加以改进,检索结果增加了题目、作者、文献等项目。总之,在 1954~1964 年计算机情报检索处于脱机检索阶段。所谓脱机检索是指用户不直接参与检索工作,由专职检索人员定期处理批量的提问要求,用户不能即刻获得检索结果。

(2) 联机检索阶段 由于在脱机检索中,用户不能与计算机进行“人一机对话”和浏览文献,就不能及时修改检索策略,检索速度较慢。所以在 1965~1972 年发展了联机检索。所谓联机检索是用户利用检索终端设备,通过通讯线路与检索系统中央计算机直接进行“人一机对话”方式的检索。可以由终端输出检索结果。联机检索灵活性大,检索者可随时浏览检索结果,若检索结果不理想,可修改检索策略,以便获得最满意的检索结果。在这一时期发展的著名的联机检索系统有:美国洛克希德(Lockheed)公司的 DIA-LOG 系统、美国系统发展公司(System Development Corporation 简称 SDC)的 ORBIT 系统、以及美国国家医学图书馆的 MEDLINE 系统等。

(3) 国际联机检索阶段 20 世纪 70 年代中后期,随着卫星通讯技术的高速发展,微型机的大量涌现以及数据库生产的迅速发展,联机检索系统打破了地域限制,实现了国际联机检索。所谓国际联机检索是指用户利用终端设备及国际卫星通讯网与世界各地的大型计算机情报检索系统的中央计算机直接进行“人一机对话”,从而查找世界各国存储在计算机系统情报文献资料。国际联机检索具有数据库内容更新快、信息量大、检索方

便、快速等特点,它的出现加速了情报资源的共享,提高了科学研究和技术开发的水平,促进了科学进步和经济发展。

我国于1980年通过香港大东电报局与美国DIALOG系统和ORBIT系统联机,利用香港这台终端开展国际联机检索业务。1983年,中国科学技术情报研究所在北京设立终端站,与DIALOG系统、ESA/IRS系统及ORBIT系统联机检索。除北京外,上海、广州、南京、重庆、长沙等几十个城市均先后设立了国际联机检索终端。

(4) 光盘检索阶段 20世纪80年代初,荷兰的飞利浦公司和日本索尼公司成功地研制出高密度光盘只读式存储器CD-ROM (Compact Disc Read-Only Memory),从而为目前光盘检索技术的普及及推广应用奠定了坚实的物质、技术基础。

光盘是一种利用激光束记录和读取信息的圆形盘片。其结构是在化学热处理过的玻璃、有机玻璃或聚碳酸酯基体上镀有铝反射层,并用碲合金膜制成信息记录层,表面覆盖一层透明聚合物起着防尘和防划痕的作用。通常应用的只读光盘直径为120mm,中心有直径为15mm的圆孔,厚度为1.2mm,质量0.15kg,其体积小,容量大,一张光盘可存储600兆字节数据,相当于1800张5.25英寸的软盘,30万页(2000字符/页)印刷品。光盘检索成本低,设备简单,在普通的微机上(包括打印机)配备光盘驱动器的驱动软件和光盘数据库的检索软件就构成了光盘检索系统。

(5) 光盘网络阶段 一张空白的光盘价值只有数元,而记录有数据的CD-ROM数据库较昂贵。例如,美国《化学文摘》(CA)光盘数据库从1987~1991年的五年累积索引,相当于第12次累积索引及其全部文摘共6张光盘数据库,价值40.5万元。为此,必须充分利用光盘数据库。在20世纪90年代,发展了光盘网络,将多张光盘数据库放入光盘塔(CD-TOWER),光盘网络上的各个检索终端可同时共享其数据库。与单机的光盘检索相比大大提高了检索效率。光盘网络化,使光盘情报服务逐渐向联机情报检索网络靠拢,光盘情报检索与联机情报检索更加紧密配合。

(6) 国际互联网(Internet)阶段 20世纪90年代以来Internet的迅速发展使得Internet正在变成各种信息资源传递的载体。它不仅向人们提供了快捷方便的通信服务,同时还是当前世界上最大的信息资源库。Internet像一张巨大的网,覆盖全球,成为名副其实的国际网络和全球信息高速公路。Internet上有成千上万台计算机,几乎每台计算机中都存储了大量信息。这些信息的内容极其丰富,按文件类型来说,有文本文件、图像文件、声音文件以及各种计算机软件程序等。可以说,Internet是人类社会有史以来第一个全球性图书馆,是一个庞大的知识信息库。作为一个科技人员,你可以通过Internet网络方便地检索世界各地的联机图书馆、数据库,访问其中的资料信息,获取有关科研项目的最新进展,同时还可以寻找同行专家,共同探讨问题,进行学术交流。

总之,只要你需要信息,就可直接进行网络检索,获取信息,为我所用。

第二节 机检系统的主要检索功能

当前一般计算机情报检索系统所执行的情报提问同文献标识或文本之间的匹配,主要有以下几种:

一、布尔逻辑检索

即采用逻辑和、逻辑乘、逻辑非等算符,将情报提问转换成逻辑表达式。计算机可以根据逻辑表达式所限定的各种运算项(检索词)的关系,确定命中文献的基本条件与查找路径,去同各篇文献的标识进行匹配。凡符合提问逻辑表达式所规定条件的文献,即作为命中文献,而予以输出。布尔逻辑符所规定的条件是:

A AND B,是指 A、B 两个检索词必须同时存在。

A OR B,是指 A 或者 B 单独一个词存在即可,或 A、B 两词同时存在。

A NOT B,是指 A 必须存在,同时不允许 B 存在。

布尔逻辑检索是把任何检索课题加工成可以进行逻辑运算的表达式,这是计算机检索的一种被广泛采用的匹配方式。

二、加权检索

所谓加权检索,就是在检索时,给每一提问检索词以一个表示其重要度的数值(即所谓“权”),然后对含有这些检索的文献词的文献进行加权计算,其和在规定的数值以上者,作为答案输出。权的大小,可以表示被检出文献的切题程度。若干命中文献按权值大小排列,这种提供情报的方式,本身就具有推荐的意义。加权检索,就是计算机在将检索词同索引词(文献标识)进行对比时,同时统计权值。然后按值的大小排序,用临界值(阈)确定输出的下限。这些工作,对于计算机来说是轻而易举的。加权检索是同布尔逻辑检索完全不同的匹配方法,但是其结果是相同的,即凡是布尔逻辑检索能实现的功能,加权检索也能实现。对于需要进行逻辑非的词,加权时可用负数。还有一种加权检索的形式,即不通过赋予权的数值和计算之和,而是在文献标引时,对不同的检索词作加权标志。对计算机来说,有加权符号的词和无加权符号的词是两个不等值的词,匹配时按不同的词进行处理就可以了。

三、截词检索

所谓截词检索,就是把检索词截断(加上截断符号),让计算机按照字的部分片断同索引词对比。这里有前方一致(右截断)、后方一致(左截断)以及中间一致(左右同时截断)三种方法。例如:

前方一致:查“计算机”(computer)这个检索词时,若用截断符号*可写成 computer*,则索引词 computer, computers, computerise, computerize 等均算命中。这种右截断的方法在计算机情报检索中被广泛使用,因为这样可以省去键入各种词尾有变化的检索词的麻烦,有助于提高查全率。当然,在何处截断,这是要认真考虑的,如果把 Computer 一词截断成 Com*,那么就会使所有以 Com 开头的索引词均被检出,其结果是误检率极大地提高。

后方一致:把截断符号放在字根的左边,如 *Computer,那么计算机进行匹配时,索引词 minicomputer, microcomputer 等均算命中。这种左截断的方式,对于某种课题的检索是很有用的。例如 *mycin,则可查出一大批有关抗菌素的文献。

中间一致:把检索词左右都同时截断,如 *Computer*,则可以命中那些其中间部分

包含这个字根的所有索引词, 如 minicomputer, microcomputer, computers, minicomputers, microcomputers 等等。显然, 这种左右同时截断的方法, 在检索较广泛课题的资料时, 能获得较高的查全率。

截词检索事实上只是计算机进行检索词与索引词之间的前方一致、后方一致、中间一致的对比匹配方式。这种匹配方式虽然手工检索时也可进行, 但计算机的效率要高得多。因此截词检索是发挥计算机本身优势、应用计算机固有的指定位对比判断功能的一种检索匹配方式, 也是计算机情报检索出现之后才产生的一种新的检索方式。

四、通用字符检索

在英语中, 有的字有不同的写法。例如“颜色”这个字有两种拼写形式 color、colour。“硫”这个字也可以分别拼写成 sulfur、sulphur。为了使不同拼写方法的“硫”或“颜色”的索引词都能被检索出来, 就出现了通用字符检索的方法。它是将拼写有变化的字母代之以通用字符符号, 如 sul * ur。这样, 计算机在进行匹配时, 凡两端的字母相同, 中间字母有所出入的都算命中, 从而使 sulphur 和 sulfur 同时都被命中。通用字符检索, 也可以叫做插入截词 (infix truncation)。事实上它是前后方一致的匹配方式, 又可称掩符检索。

五、字符串检索和相邻度检索

所谓字符串, 是指字母、空格及规格规定的其他可能的符号的序列。字符串检索, 就是使用位置逻辑检索的功能, 规定检索的目标必须出现一些什么字符, 各字段之间距离应不超过多少字符位置, 以及前后位置能否进行互换。字符串检索主要用于二次检索。即在已检索到一定的文献集合的基础上, 用它来作精细检索。

首先, 字符串检索可用来查找和排除一定的字。例如, 关于“铬”的一些名词是:

Chromates	铬酸盐
Chromating	铬酸盐处理
Chromatography	色谱
Chromic acid	铬酸
Chromite	铬铁矿
Chromizing	铬化
Chromium	铬

以上这些字, 除“色谱”外, 全是同“铬”有关的。如果在检索时, 想命中这些同“铬”有关的索引词, 并且要排除“色谱”的话, 那么光靠上面说过的截词检索和通用字符检索是不行的。因为这两种方法都不能在保留所有“铬”家族的词的同时, 将“色谱”这个“异己”的词排除掉。这就需要另一种匹配方法即字符串检索。

字符串检索的表示方法很多, 其中如:

Chrom I A 0 atography

这里, 中间的“I”表示符合此条件的否定。“A”表示前后两个字符串不能对换, “0”即为前后两个字符串彼此间隔为零。这样, 就把“Chromatography”(色谱)的文献排除了。

其次,相邻度检索可用来检索一定的词组。所谓词组,又可称多元词,即由几个固定次序的字所组成的词,例如:“invitro toxicity testing”(试管内的毒性试验),或“Communication Satellites”(通讯卫星)等等。对于词组,如果用单字进行逻辑乘的组配运算,检索出来的不一定能完全符合要求。这时就得用字符串检索功能了。例如可以规定“通讯”在前,“卫星”在后,两者紧密相邻,前后不能互换等等的位置逻辑。当然有的前后可以对换,如“虐待”与“儿童”两者用相应的符号(依系统规定为准,如“S”)。此外,规定各个字符段之间相隔一定的距离,也是必要的,例如查找“情报检索”,“情报”与“检索”之间允许插入一些字,如“情报的存储与检索”。

相邻度检索在 Dialog 等其他系统中,采用不同的标记形式。如有词 A 及词 B 两个词,若指定它们必须出现在同一篇文献中,则用 A (C) B 表示;若指定它们必须在同一字段出现,则用 A (F) B 表示;若指定该二词中间可相隔若干(n) 个词,则表示为 A (nW) B。如书名 GONE WITH THE WIND (飘),可写成 GONE (2W) WIND 这样的表达式。在这里,“W”是“With”的缩写。它的意思是规定此符号的前后字符不能互换位置。另外还有一些位置逻辑的算符:

“N”是“near”的缩写,表示此符前后的字符可以互换位置,但必须紧密相邻,即相邻度为 0。当然,“N”前也可标明系数,指明相邻度的值,以规定中间最多可插入多少个字。“nN”与“nW”的用法相似。

“L”是“link”缩写,表示两侧的字符具有一定的从属关系。这种字符主要用于一级主题词与二级主题词的相连关系。例如“钢(L)连铸”,是指“钢”为一级主题词,“连铸”为二级主题词。

第三节 检索策略

当前,我国有些计算机情报检索系统已投入使用,向情报用户提供检索服务,并且还利用联机终端,对国外的检索系统进行检索。有关检索策略的研究,日益受到人们的重视。检索策略构造的好坏,直接影响到相关文献的查全率和查准率,关系到检索服务的效果。当然,对于一定的情报提问来说,检索效果的优劣取决于许多因素:首先是书目数据库本身的因素,如该数据库所使用的词表质量、标引质量、文献摘储率等;其次是系统软件所能提供的功能。但是,在一定的数据库质量与系统功能的前提下,检索策略无疑是一个非常重要的因素。在同一个检索系统中对同一书目数据库进行同一课题的检索,不同的检索策略会导致不同的检索效果。对于计算机情报检索来说,检索策略的构造与调整,是一个检索服务人员可以施展其技巧本领的广阔“舞台”。

由于大多数计算机情报检索系统一般都采用控制词汇和布尔逻辑组配的检索方式,因此我们把重点放在控制词汇条件下的布尔检索策略的构造与调整方法上。

一、检索策略的构造步骤与方法

构造一个检索策略,可遵循下列步骤:

1. 弄清楚用户的提问要求,并确定所查找文献在类型、文种、时间等方面的范围;
2. 根据上述要求选择数据库(或文档)并确定查找途径;

3. 对用户的提问进行概念分析, 按照检索词表, 选择能代表各个组面的检索词;
4. 运用逻辑运算符构造检索表达式;
5. 对文档进行查找, 检出相关文献并分析检索结果;
6. 如有必要, 以提问式进行反馈修改, 并且重复第 5 步, 直到用户满意为止。

在构造检索策略过程中, 首先必须搞清楚用户想查找的是关于什么主题的、哪些类型的文献、以及他的情报需求范围。分析检索课题的实质时, 应将情报提问按内容实质分解成若干个不同的概念组面。

其次, 必须把各组面概念转换成词表中的词。任何情报提问的概念分析的完成, 依赖于检索语言精确表达情报提问所涉及概念的能力, 即词表的网罗度与专指度。因而在把概念转化成检索词时, 要求从词表中选择最合适的词来表达, 或者在构造检索策略时利用叙词表的参照结构和等级关系进行选词。选词时不仅要从字面上拆词, 更主要的是从词的含义上进行拆义。要注意到词表中主题词的专指深度和主题词的漏选情况。还需要注意下列选择检索词的原则:

1. 要从词表规定的专业范围出发, 选用各学科内具有检索意义的基本名词术语;
2. 避免选择使用频率低的词;
3. 应多选用基本词汇进行组配;
4. 一般不选用动词和形容词。

选定检索词之后, 就可以用布尔关系运算符如逻辑或 (OR)、逻辑与 (AND) 来表达用户的情报提问中各有关概念之间的关系, 并可利用括号来体现各检索词之间的完整性和运算的完整性及运算的优先顺序。提问式有时很难做到一次定准, 往往需要经过几次“上机检索—分析结果—修改提问”的试验过程。此外, 在编写逻辑提问式时, 要注意下列技巧:

(1) 对用 AND 连接的检索词, 应把出现频率不高的放在 AND 的左端, 这样, 如果出现否定的回答, 则能最快地出现, 以节省计算机处理时间。

(2) 对用 OR 连接的检索词, 应把出现频率高的词放在 OR 的左端, 这样, 如果出现肯定的回答, 则能最快地出现, 以节省计算机处理时间。

(3) 当 AND 和 OR 两种运算混合出现在一个提问式中时, 如果连续有几个 OR 运算出现, 则应放在 AND 的左边。

二、检索策略的调整技术

检索策略, 具体来说是提问逻辑式, 对于查全率与查准率有极大的关系。在控制查全率与查准率方面, 有若干调整的措施。

去掉用“与”逻辑 (AND) 连接的非主题限定词 (例如文献类型、出版年代、文种等) 以及增加用“或”逻辑 (OR) 连接的相关检索词, 这两种方法对于提高查全率的效果比较直观明显。删除检索式的某个组面, 可通过降低检索式的网罗度而扩大检索策略的范围, 以达到提高查全率的目的。但同时可能检出一些与用户情报需求无关的文献, 从而降低了查准率。

如果说标引的网罗度对查全率的影响较大, 则检索词的专指度对查准率的影响更大。检索词的专指度就是指检索词表的适用性及其揭示文献主题的深度。在构造检索策略时,

可用提高检索词的专指度的方法来改善查准率。例如,有些太泛指的词,如“分析”,应建议使用较专指的词,如光谱分析、化学分析、数学分析、物理分析以及计算机分析等。

大多数检索课题的特点是高专指性的,而不是泛指的。专指性的内容多数情况下是由课题中各方面的因素来决定的。一般来说,当用户感兴趣的精确课题没有被词表中的合适的专门词所包括时,放宽检索策略可以说是合理的;而当此课题已被专门词包括时,放宽检索策略则是不合理的了。从查准率的观念来看,把提问的两个组面同时进行族化,通常会造成严重的误检。

其他的检索方法,如:同义词控制、聚类法、截词法以及用于提高查准率的几种检索方法,将在以后叙述。

在检索过程中,往往利用文献的外表特征(如出版年代、文献类型等)在检索表达式中通过“与”逻辑增加一个限制因素,便可以大大减少检出的文献量,提高查准率。要实现检索策略所期望的接近相关文献、淘汰无关文献及扩大检索成果的目标,大致有三种方法:

1. 积木法。把检索课题分解成若干组面。先就每一个组面通过选用若干个用“或”逻辑相连的词进行检索,分别得出初步结果,然后各个组面的结果再用“与”逻辑组配起来。这种积木法的优点是能够比较清楚地提供检索逻辑史,以便以后去检索与回顾,并保留与调用检索式。它的缺点是:要耗费较多的联机时间。

2. 引文珠串增长法。这是用每个概念组面的最专指的词开始直接检索,以便找到少量的对口引文。然后联机浏览一篇或几篇引文,从中找出更多的索引词和自由文本,并将其纳入检索式,从而可检索到更多的引文。这种扩大成果的方法的优点在于,用动态的、实验的方式去得出较优的检索结果。这种方法用来浏览引文与思考选词的联机时间较多,但对选择检索词却很有帮助。ESA—IRS/系统的 ZOOM 功能有助于珠串增长法的实现。

3. 逐次分割法。即先建立一个较粗泛的集合,然后用其他必要的组面通过“与”逻辑进行缩小集合,像是削笋一样,最后获得相关引文的核心部分。

这种方法是一个由粗到精、逐步提炼的过程。命中文献数量逐步减少,直到获得所期望的结果。在检索过程中,可以随时根据命中的文献量确定检索的专指度。

以上说明了调整检索策略所用的各种方法及其对查全率和查准率的影响。但需要指出的是,检索策略的确切形成要受到所查找的数据库特性的影响。弄清所使用的数据库以及整个检索系统的基本性能,是制定一个良好检索策略的前提条件;熟悉检索系统操作指令的基本功能和文献检索的组配理论方法,是制定一个恰当的检索表达式的基础;而同用户进行反复的商量讨论,准确地表达出用户情报提问的实质及需求范围,则是构造合理的检索策略的关键。

联机检索系统可以通过人机交互来实现检索策略的改进和输出情报的最佳化,脱机的定题情报服务同样可以通过系统与用户之间的反馈联系来提高查全率和查准率。可以采用的一种方法是:在为用户寄送检索结果时就附上一份征求意见的评价单,在评价单上列出所使用的检索词及逻辑表达式,要求用户对检索结果是否符合情报需求以及有无误检与漏检等,提出评价意见寄回来。这种评价反馈可以定量地判断查全率和查准率,有助于深刻掌握用户的情报需求,准确地构造检索策略,从而取得较好的检索效率和情报服务效果。

检索策略的优劣程度主要取决于检索者的水平。具体说来是:

1. 检索者弄清提问实质并准确进行概念分析的能力;
2. 检索者选择合适的词用以表达所要查找主题内容的能力;
3. 检索者把选出的词以逻辑上的“完整性”结合在一起的能力;
4. 检索者设想所有合理的检索途径的能力;
5. 检索者通过改变提问表达式的网罗度和专指度等一系列调整技术,以满足用户对查全率和查准率要求的能力;
6. 检索者检查检索的效果,并以此作为“反馈”不断修改检索策略的能力。在实际的检索中,往往产生一些失误,即漏检和误检。也就是说,该找的文献没有找全,无关的文献倒被查出来了。克服检索失误,提高查全与查准性能,应该说是联机检索服务的生命线。

造成检索失误的原因是多方面的。如标引质量、词表质量、检索程序的功能、系统与用户的交互功能、检索策略的构造质量、数据库的选择等等,都是相关的因素。

第二章 DIALOG 查询系统

第一节 DIALOG 系统简介

我国的联机检索服务从 20 世纪 80 年代开始,通过香港大东电报局与美国 DIALOG 系统和 ORBIT 系统联机,利用香港这台终端开展国际联机检索业务。

除北京外,我国已在上海、广州、南京、重庆、长沙等 50 多个城市先后设立了一百多个国际联机终端,通过卫星通信线路和十多个国际性大型信息检索系统联机,它们是: DIALOG、ORBIT、BRS、ESA-IRS、STN、DMS/DRI、ECHO、INIS、Pergamon-Info-Line、WRS、MEDLINE、Data-Star 和 OCLC 等。

中国国防科技信息中心于 1981 年建立了中国第一个计算机联机信息检索系统,使用自行设计、研制的信息检索软件 BDSIRS。目前已拥有九十多个专线联机终端,可向全国 20 个省市提供服务。其数据库包括:美国政府报告库、英国科学文摘库、国外期刊库和中国专利库、国防科技报告库、国内成果交流库、中外标准库等十余种。

中国科技信息研究中心,是我国最大的一个科技信息中心,它和机电部情报所、化工部情报所联网后形成整体优势,向全国联机终端提供检索服务。许多数据库是我国自建的,如:中国学术会议论文库、中国学位论文库、中国科技成果库、中国专利库、中国科技期刊联合目录库、中国国家标准文献库、中国企业公司及产品库、中国机械工程文摘库和中国化工文摘库等等。

我国业已开始了光盘数据库的生产,有些光盘也将提供联机检索。科技、学术性的光盘版数据库有中国科学院重庆分院出版的《中国科技期刊论文篇名数据库》,清华大学出版社出版的《中国学术期刊(光盘版)》全文数据库,中国人民大学书报中心出版的《中国报刊论文数据库》等等。《中国专利文献数据库》光盘也已由中国专利文献出版社发行,《中国大百科全书》74 卷的光盘已于 1996 年底全部完成。现在全世界有一百多个联机检索系统,目前在我国用得最多的有 DIALOG 系统、OCLC 系统和 STN 系统。这些也就是本部分着力向读者介绍的内容。

美国 DIALOG 情报服务公司拥有当今世界上最大的联机情报检索系统,总部设在美国加利福尼亚旧金山的帕洛·阿尔托市(Palo Alto)。该系统从 1972 年起服务至 1988 年 6 月已有三百多个数据库,专业内容有:社会科学、化学、物理、数学、医学、材料学、时事、建筑学、教育、经济学、商业、公司、人文学、宇航、能源、法律、工程、农业、计算机、生物等。数据库文献类型有:杂志、会议录、图书、论文、进展中科技报告、政府文件、专利、标准、市场行情、经济预测、统计数据、名人传记等。

DIALOG 数据库由数据库供应商(主要由私人、国家或政府机构提供的出版物)提